A ROBUST SCHEME FOR OBLIVIOUS DETECTION OF WATERMARKS / DATA HIDING IN STILL IMAGES *

Mahalingam Ramkumar and Ali N Akansu
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
New Jersey Center for Multimedia Research
University Heights, Newark, NJ, 07102.

ABSTRACT

We propose a novel, robust scheme for data hiding/ *oblivious detection* of watermarks in still images. While the low-frequency image coefficients are *robust*, they cannot be used effectively for oblivious detection methods, when *correlative processing* is employed for detection. However, in the proposed non-linear detection method, the robust low-frequency bands can be used effectively. Thus the proposed method turns out to be more robust than methods employing linear addition and correlative extraction of the signature. We report the results obtained for 7 test images in terms of probability of error in detection of the watermark/ hidden bits.

Keywords: W atermarking, Data Hiding, Oblivious Detection.

1. INTRODUCTION

Watermarking is a means of embedding a hidden 'signature' in multimedia digital data for copyright protection. The vulnerability of digital data to intellectual piracy without loss of quality has become a major concern with the advances in storage and copying technology. This has spurred significant activity to look for ways and means to detect and trace copyright violations. In this paper, we restrict ourselves to watermarking of still-images.

From the standpoint of communication theory, invisible watermarking, (synonymous to data-hiding, invisible communications, and steganography) is a form of very low bit rate communication (the amount of information conveyed by the watermark, in number of bits, is very small when compared to that of the original image) under low signal to noise ratio. Naturally one looks towards spread-spectrum communications to achieve this. Given a *cover* image I the watermark takes the form a spread-spectrum signal S added to the image, where S in general is dependent on I.

$$\hat{I} = I + S, \tag{1.1}$$

is called the *watermarked* image. The originator of the cover image I inserts the watermark S before it is sold to the buyer as $\hat{I}$.

Consider the following scenario, in which the buyer of $\hat{I}$, creates an (unauthorized) image $\tilde{I}$ by processing $\hat{I}$ such that

$$\tilde{I} = \hat{I} + N = I + S + N \tag{1.2}$$

where N is the 'noise' introduced due to processing. In general

$$\tilde{I} = \mathcal{T}(\hat{I}) + N = \mathcal{T}(I) + \mathcal{T}(S) + N \tag{1.3}$$

where $\mathcal{T}$ is a transformation that may involve translation, rotation, scaling, and cropping. By comparison of the images I and $\tilde{I}$, however, one can obtain a reasonable estimate of the transformation $\mathcal{T}$, which can then be approximately inverted (except of course, the cropped part of the image - a normal way of dealing with that is to replace the cropped parts with portions from the original image). After the inverse transformation, we could now, without any loss of generality use the form of Eq.(1.2) as the mathematical model for the image $\tilde{I}$.

*This work was partially supported by Panasonic Technologies Inc., Plainsboro, NJ.

The originator of I would now like to prove in a court of law that the image $\tilde{I}$ is an unauthorized reproduction of I. In addition he/she would also like to establish the identity of the *pirate* responsible for unauthorized circulation of some form of the image $\hat{I}$. The originator should be able to do all of the above by extracting information from the watermark in $\tilde{I}$. The spread spectrum signal S, should therefore contain the pertinent bits of information - viz., the identity of the creator and identity of the buyer (who is the *pirate* in this case) and possibly other useful indexing information.

The obvious way of extracting S from $\tilde{I}$, when the original image I is readily available is to subtract I from $\tilde{I}$ and get the residual $S + N$. The problem is now of extracting S from $S + N$. Such a scheme, involving modulation of the DCT coefficients can be found in Cox *et. al.*[1] However, this method of extracting the watermark by *direct involvement* of the original image (such schemes are termed as *cover image escrow* hiding schemes) poses some legal problems[2,3] in establishing the originator of I. As the watermark has been extracted by subtracting I from $\tilde{I}$, the pirate could equally well claim that $\tilde{I}$ is the original image and I has been created by adding a watermark $\hat{S} = -S$ to the "original" image $\tilde{I}$.

A way out of predicament is to extract the watermark *without direct involvement* of the original image I (such schemes are termed as *oblivious detection schemes*). This scheme would then boil down to extracting S from $\tilde{I} = S + (I + N) = S + N_1$, where, $N_1 = I + N$ is in general much greater than N. This would therefore necessitate S of much greater energy to effectively communicate as many bits as the scheme in which the original image is subtracted from $\tilde{I}$.

Watermarks are usually detected by correlative processing of the spread-spectrum sequence (the watermark) with the image (or its representation in some domain) in which the presence or absence of the signature is to be tested. If the result of the correlation is above a certain threshold it is decided that the signature is present. Under this circumstance it is very easy for the pirate to 'design' his own watermark (or signature) $\hat{S}$ which yields a high correlation with $\tilde{I}$. As $\tilde{I}$ and I are *statistically similar*, $\hat{S}$ would also yield a high correlation with I. So while the true originator can show his signature S in both the images, so can the pirate. The only way to avoid this dead-lock is to have rigid guidelines on how an authentic signature may be obtained, as for instance, in Zeng *et. al.*[3]

Note that the complete signature need not be detected obliviously. Once the originator's identity is detected obliviously, the other parts of the watermark (like identity of the buyer) can be detected with direct involvement of the original (unwatermarked) image. Detection of the originator's identity is equivalent to detecting a single bit, which answers the question "Is the signature of the originator present in the image?"

2. CHANNEL FOR HIDDEN COMMUNICATIONS

Now consider the problem of a secret communication channel between a transmitter and a receiver, in which a few bits of information are hidden in an image such that the image is not visibly altered. The image may undergo some modifications *en route* to the receiver, like lossy compression. Furthermore, the receiver has no information about the specific image which is the carrier of secret information (the bits are to be detected obliviously).

Figure 1 (a) is a block diagram of a hidden communications channel. $\boldsymbol{S}$ is the message to be transmitted through the channel which has two sources of noise; $\boldsymbol{I}$, the noise due to the cover image, and $\boldsymbol{P}$, the noise due to processing (compression / decompression). We could decompose the image $\boldsymbol{I}$ into multiple (frequency) bands, and transmit $\boldsymbol{S}$ through multiple channels, with different characteristics. Figure 1 (b) illustrates the decomposition of the communication channel into L channels, corresponding to decomposition of the image into L bands. The signature energy is also split into components $S_1 \cdots S_L$. We know that typically, the low frequency channels have a large amount of cover image noise. On the other hand, a high amount of noise is likely to be introduced in the high frequencies due to processing of the image (like lossy compression).

Apparently, very low and very high frequency bins should not be used for the communication scheme. The problem of detecting a watermark (obliviously) can be considered as a special case of a hidden communication scheme, where the presence or absence of *one* bit is to be ascertained. Note that the main difference between the cover image escrow schemes and oblivious detection schemes is that the noise in the channel is only the processing noise for the former case. So cover image escrow schemes should concentrate the signature energy in the low frequency bands (bands for which processing noise is low). On the other hand, oblivious detection schemes should utilize the mid-frequency bands. So an efficient watermarking scheme would use the mid-frequency bins for hiding the signature corresponding to the originator's identity (which is to be detected obliviously), and the low frequency bins to hide the remaining

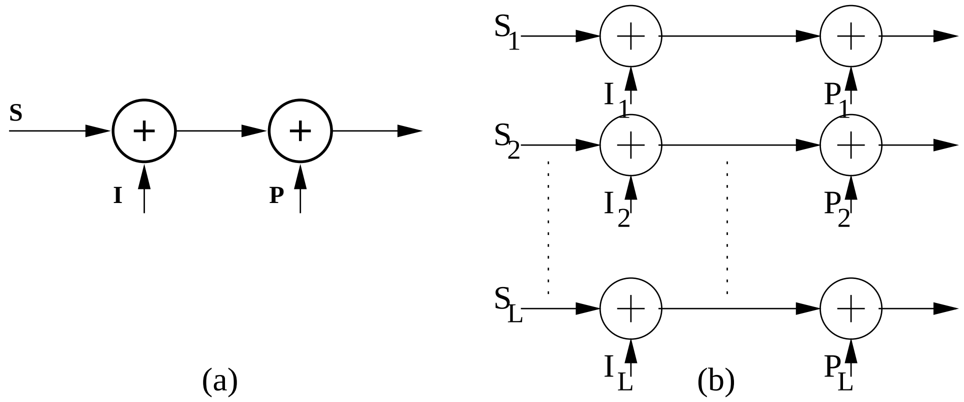

Figure 1. (a) Channel for Hidden Communication. (b) Decomposition of the channel of (a) into multiple channels

components of the signature (buyer's identity etc.). However, using only the mid-frequency bands may imply reduced robustness.

3. EXISTING METHODS

In this section we shall briefly review some of the proposed methods that permit oblivious detection. In most methods, the first step is *image noise reduction*, in which a large portion of the image energy is made orthogonal to the signature. This can be done either by neglecting the low frequency coefficients of a frequency transform of the image like DCT or DFT (in Zeng *et. al*[3] for instance, the lowest frequency (dc) DCT coefficient of each 8×8 block of the image is neglected). Another way might be to employ prediction of a pixel, from its neighboring pixels. The signature is then added to the prediction error, which has much lower energy than the original image. This method for example is adopted in Kutter *et. al.*[5]

Once a substantial portion of the image energy (noise) is subtracted from the original image the signature is added to the residual, or part of the residual. The main disadvantage of the method in Kutter *et. al*[5] is that the signature is added to the entire residual image, regardless of the distribution of the signature in the frequency domain. As a result, this method is not very robust for lossy compression. The signature addition is performed by forcing many residual coefficients to have one sign rather than the other. Detection is performed by counting the difference between the number of positive and negative coefficients. A major advantage of this scheme is that the signature would be unaffected by histogram equalization. In Zeng *et. al*,[3] the signature takes the form of a Gaussian random sequence scaled by a visual threshold value corresponding to each frequency bin. The detection statistic is a measure of correlation between the signature and the DCT coefficients of the signed image. The measure of effectiveness for a watermarking scheme is the separation that can be achieved between the statistics obtained from an unsigned image (or when the presence of a wrong signature is checked for in a signed image) and a signed image. The separation is a measure of confidence in detection of the signature.

4. THE PROPOSED METHOD

The main problem with oblivious detection schemes is that, suppression of image noise causes one to loose the bands that are most robust to processing. We therefore propose a non-linear method which permits use of low-frequency, high image noise bands and yet achieves image noise suppression. This method therefore, turns out to be more robust than previously proposed methods. This method permits detection of the signature with a very high degree of certainty even when the image has gone through JPEG compression with a quality factor of 15. The watermark is also detected with a low probability of error in images even after Printing-reScanning (PS) and Printing-Photocopying-reScanning (PPS).

4.1. Embedding the Watermark

A block diagram of the watermark embedding scheme is displayed in Figure 2. In this method the signature is embedded in the DCT domain. The 8×8 DCT of the 1024 blocks of the image is performed. 9 DCT coefficients

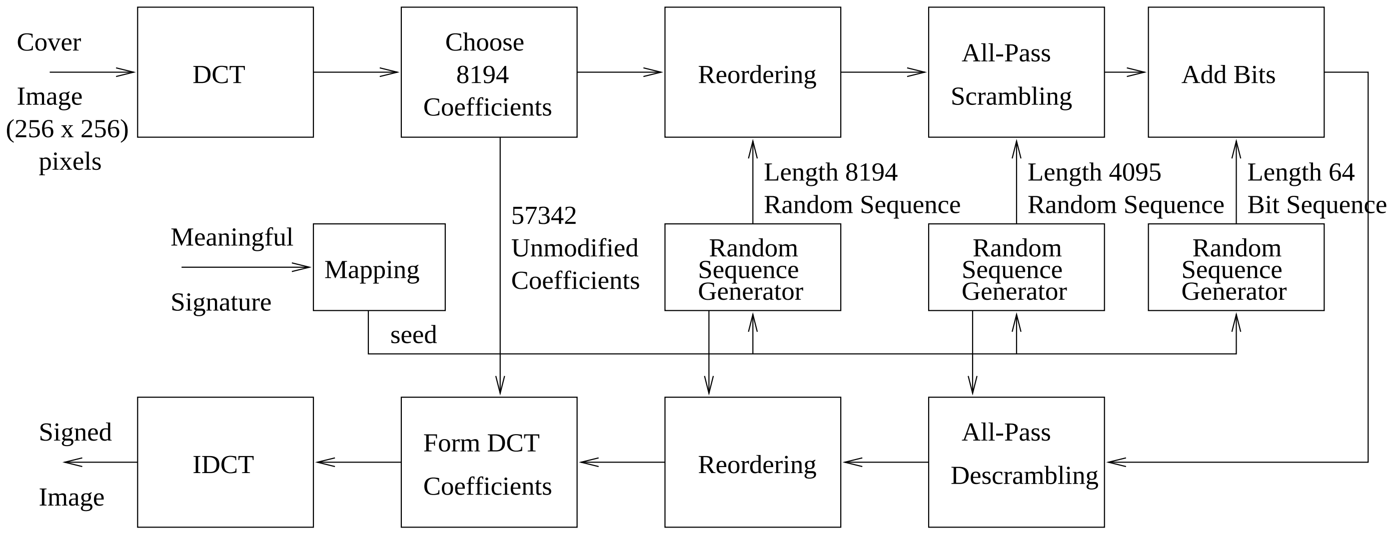

Figure 2. Watermark Addition

from each block are arranged as a vector by zig-zag scanning. The first coefficient (the DC coefficient) is neglected, and the remaining 8 low-frequency coefficients of each block are stacked to form a vector $\boldsymbol{v}$ of length 8192 (8×1024). A *meaningful signature* is mapped to a seed which is used to generate a uniform random sequence of length 8192, which in turn is used to generate an ordering sequence $\boldsymbol{o}$. The ordering sequence $\boldsymbol{o}$ is used to reorder the sequence $\boldsymbol{v}$ to obtain a

$$\boldsymbol{v}_r = \boldsymbol{v}(\boldsymbol{o}). \tag{4.4}$$

In the next step, we generate another uniform random sequence, $\boldsymbol{\Theta}$ (between π and $-\pi$) of length 4095, with the same seed. These values are used as the angles of a length 8192 cyclic all-pass filter $\boldsymbol{A}$, which is used to scramble the vector $\boldsymbol{v}_r$.

$$\boldsymbol{v}_{\text{ra}} = \text{idft}(\text{dft}(\boldsymbol{v}_r).\boldsymbol{A}) \tag{4.5}$$

where

$$\begin{aligned} |A(m)| &= 1,\, m = 1 \cdots 8192 \\ \angle A(m) &= 0,\, m = 1,\, m = 4097. \\ &= \Theta(m),\, m = 2 \cdots 4096 \\ &= -\Theta(8192 - m + 2),\, m = 4098 \cdots 8192. \end{aligned} \tag{4.6}$$

We now embed 64 bits corresponding to a signature in $\boldsymbol{v}_{\text{ra}}$, where each bit is embedded in 128 coefficients of $\boldsymbol{v}_{\text{ra}}$. Let $\boldsymbol{v}^i_{\text{ra}}$ be the 128 coefficients corresponding to the bit i. If $i = 1$, we then modify all coefficients of $\boldsymbol{v}^i_{\text{ra}}$ that have values between $-t$ and 0 by adding a positive number to them so that the coefficients end up with a positive sign. If $i = 0$, we modify all coefficients with values between 0 and $+t$ by adding a negative number. Let $\boldsymbol{v}^i_{\text{ras}}$ be the modified coefficients and $\boldsymbol{v}_{\text{ras}}$ the corresponding vector of length 8192. $\boldsymbol{v}_{\text{ras}}$ is now unscrambled by inverse all-pass filtering to obtain

$$\boldsymbol{v}_{\text{rs}} = \text{idft}(\text{dft}(\boldsymbol{v}_{\text{ras}}).\text{conj}(\boldsymbol{A})). \tag{4.7}$$

$\boldsymbol{v}_{\text{rs}}$ is then re-ordered to get $\boldsymbol{v}_s$, where

$$\boldsymbol{v}_s(\boldsymbol{o}) = \boldsymbol{v}_{\text{rs}}. \tag{4.8}$$

$\boldsymbol{v}_s$ is thus $\boldsymbol{v}$ modified to add the 64 bits (or the watermark). We substitute $\boldsymbol{v}_s$ for $\boldsymbol{v}$ in the DCT coefficients of the original image and take the inverse DCT to get the watermarked image.

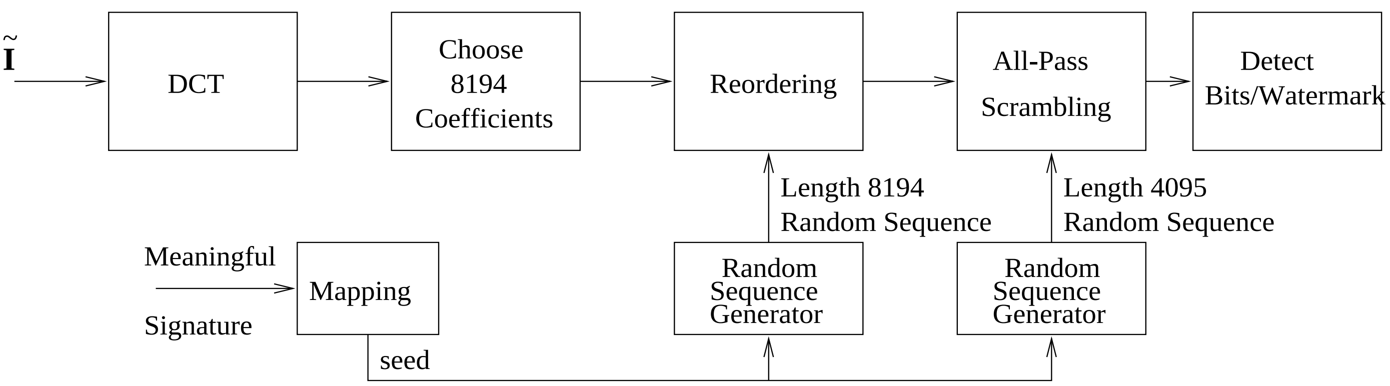

Figure 3. Watermark Detection

Image	PSNR	NOP	J-25	J-15	CR	SS-4	SP	B30	B25	B15
Lena	41.76	1732	1128	726	702	786	224	0.01	1.12	14.8
Barbara	41.39	1934	1424	1018	856	958	354	0.00	0.34	5.4
Goldhill	40.34	2508	1892	1536	1118	1024	632	0.00	0.46	3.2
Boats	41.12	2020	1362	988	852	894	396	0.00	0.56	6.0
Baboon	40.15	2468	1826	1382	936	866	344	0.00	0.18	4.4
Peppers	41.61	1882	1326	922	716	920	438	0.01	0.98	6.6
Airplane	41.94	1834	1266	874	838	726	454	0.02	1.16	6.8

Table 1. The Test Statistic q and Bit Error Rates Obtained from Watermarked Images After JPEG, SPIHT, Cropping, Sub-sampling.

4.2. Detecting the Watermark

A block diagram of the watermark detection algorithm is shown in Figure 3. The 8×8 DCT of the image in which the watermark is to be detected is taken and the 8192 coefficients $\tilde{\boldsymbol{v}}$ are obtained. The ordering sequence $\boldsymbol{o}$ and the angles of the all-pass scrambling sequence are generated from the known seed. $\tilde{\boldsymbol{v}}_{\text{ras}}$ is obtained by reordering and all-pass scrambling of $\tilde{\boldsymbol{v}}$. The 128 coefficients $\tilde{\boldsymbol{v}}^i_{\text{ra}}$ corresponding the bit i is obtained. Let $s_i(m) = \pm 1$, $m = 1 \cdots 128$, represent the signs of the 128 coefficients of $\tilde{\boldsymbol{v}}^i_{\text{ra}}$. Let

$$S(i) = \sum_{m=1}^{128} s_i(m). \tag{4.9}$$

If $S(i) > 0$ the embedded bit is 1. If $S(i) < 0$ the embedded bit is zero. For the purpose of detecting the presence of a watermark (corresponding to the originator's identity, where only one bit has to be detected), if $\boldsymbol{b}$ is the embedded bit sequence, we obtain the statistic

$$q = \boldsymbol{b}\boldsymbol{S}^T \tag{4.10}$$

where, q is a measure of the certainty with which the signature is detected.

4.3. The Statistic q

Over one million samples of the statistic q were obtained by checking for a signature in unsigned images, or by checking for a wrong signature in signed images. The histogram of the values are plotted in Figure 4. It is seen that the random variable q is very close to a Gaussian distribution with standard deviation 90 (dotted curve).

The probability of error in detecting a signature, when $q = 270, 360, 450, 540$, (obtained from Q-Tables) are correspondingly $Q(3) = 1.35 \times 10^{-3}$, $Q(4) = 3.17 \times 10^{-5}$, $Q(5) = 2.87 \times 10^{-7}$ and $Q(6) = 9.87 \times 10^{-10}$ respectively.

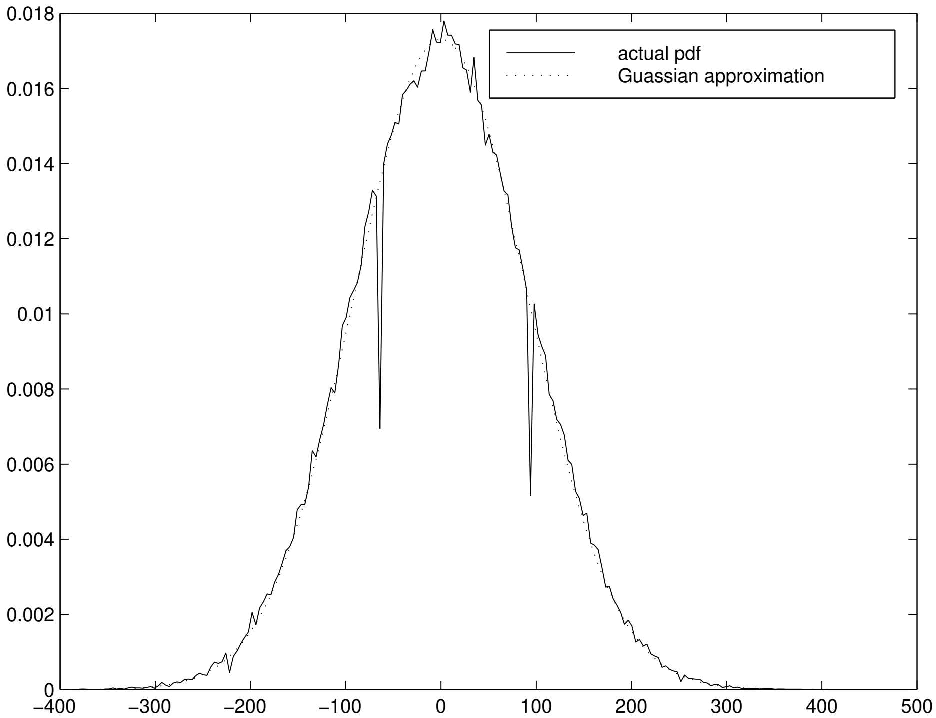

Figure 4. The distribution of the test statistic q

Figure 5. 256×256 Original Barbara image

Figure 6. Watermarked Barbara image

5. PERFORMANCE EVALUATION

Figure 5 shows the original Barbara image, and Figure 6, the visually identical watermarked image (PSNR 41.39 dB). Results of the performance of the proposed scheme for seven 256×256 test images are reported in Table 1. The PSNR of the signed image and the values of the statistic q obtained after watermarking with no processing noise (NOP), JPEG compression with quality of 25 (J-25) and 15 (J-15), cropping (CR) 50% of the pixels of the image, sub-sampling (SS-4) by a factor of 4, and SPIHT (SP)[6] compression at 0.25 bpp are tabulated. It is seen that the probability of error (P_e) in detection of the signature after JPEG-15, is 3.6×10^{-16} for Lena (corresponding to $q = 726$) and 1.3×10^{-65} for Goldhill image (corresponding) to $q = 1536$). Also listed is the average number of bits in error (out of 64) for JPEG with quality 30 (B30), 25 (B25) and 15 (B15). The bit error rates were averaged over 100 different signatures.

Further the signed and unsigned images were subjected to the PS and PPS cycles. q values of 587 ($P_e = 3.4 \times 10^{-11}$) and 432 ($P_e = 8 \times 10^{-7}$) were obtained for PS and PPS watermarked Barbara image. The corresponding values for the unwatermarked PS and PPS images were -87 and -12 respectively. The PPS Barbara image is shown in Figure 7. It is also seen that on an average, less than 1 out of 64 bits were in error following JPEG-25. However, after JPEG-30 there is virtually no error.

6. CONCLUSIONS AND FURTHER WORK

The main motivation for the proposed scheme comes from the observation that non-linear methods for detection of the signature (like sign-counting) seemed to perform better (in terms of the separation of the statistic q obtained for signed and unsigned images), than linear addition and correlative detection of the signature. The improvement comes from the fact that this method (unlike methods that use correlative processing) facilitates an alternate method of image noise suppression, thereby permitting us to use the robust low frequency bands. The image noise suppression is done by ignoring 'large coefficients' after all-pass filtering. On the other hand for effective detection by correlative processing, it is necessary to suppress the robust low frequency bands completely. Thus this method is equally effective even if the DC coefficients are used. But the reason for leaving the DC coefficients intact is that modifying them may lead to unacceptable blockiness.

Figure 7. 256×256 Barbara image after Printing - Photocopying - reScanning cycle

The main disadvantage of this method is that, unlike linear methods, it is very difficult to spread the energy of the signature optimally among different coefficients based on their visual thresholds. All-pass scrambling tends to distribute the signature energy uniformly amongst all the 8192 coefficients, even when only a few coefficients in the all-pass scrambled domain ($\boldsymbol{v}_{ra}$) are altered. As of now we solve this problem by just adding larger signature energy initially and then chopping off some signature components in the DCT domain to ensure that they do not exceed the visual thresholds. Though clearly suboptimal, the method still performs better than any of the existing methods. Presently we are trying to obtain a recursive algorithm for maximizing the number of coefficients whose signs are altered in the all-pass scrambled domain, while maintaining the desired distribution of the signature in the DCT domain.

REFERENCES

1. I.J. Cox, J. Kilian, F.T. Leighton, and T.G. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", IEEE Transactions on Image Processing **6** (12):1673-1687, 1997.
2. S. Craver, N. Memon, B-L. Yeo, M. Yeung, "Can Invisible Watermarks Resolve Rightful Ownerships", IS & T/ SPIE Electronic Imaging: Human Vision and Electronic Imaging, Vol **3022**, pp 310-321, Feb. 1997.
3. W. Zeng, B. Liu, "On Resolving Rightful Ownerships of Digital Images by Invisible Watermarks", Proceedings of ICIP, vol **1**, pp 552-555, 1997.
4. T. M. Cover, J. A. Thomas, *"Elements of Information Theory"*, Second Edition, John-Wiley and Sons Inc., 1991.
5. M. Kutter, F. Jordan, F. Bossen, "Digital Signature of Color Images using Amplitude Modulation", Proceedings of SPIE-EI 97 Storage and Retrieval for Image and Video Databases, San Jose, LA, 3022-5, pp 518-526, Feb. 13-14 1997.
6. A.Said, W.A.Pearlman, "A New Fast and Efficient Implementation of an Image Codec Based on Set Partitioning in Hierarchical Trees", IEEE Transactions on Circuits and Systems for Video Technology, Volume **6**, pp. 243-250, June 1996.